



CERES Data Management Group Status and ASDC Status Report

CERES Science Team Meeting

October 7th, 2014

Jonathan Gleason - CERES DMT,
John Kusterer - ASDC,
Lindsay Parker - SSAI – ASDC



Data Management Team

**Systems/
Optimization:**
Nelson Hillyer
Michael Linsinbigler
Josh Wilkins

**Production /
Optimization:**
Lisa Coleman
Carla Grune

Instrument:
Denise Cooper
Thomas Grepotis
Dianne Snyder

Convolution:
Igor Antropov

Inversion:
Victor Sothcott

**Product Web
Support:**
Churngwei Chu
Ed Kizer
Cristian Mitrescu

**Configuration
Management:**
Tammy Ayers
Joanne Saunders

ERBELike:
Dale Walikainen
Jeremie Lande

SARB:
Tom Caldwell

Clouds:
Sunny Sun-Mack
Ricky Brown
Yan Chen
Liz Heckert
Rita Smith
Sharon Gibson

TISA Gridding:
Raja Raju

Task Manager:
Walt Miller

TISA Averaging:
Cathy Nguyen
Dennis Keyes



Overview

Data Management Status

Current Data Availability

Recent Activities (Deliveries)

Automation Effort Status

Software Delivery & Data Release Milestones

MODIS C5 / C6 availability status

System Migration Update

ASDC Status

Upcoming Outages

Big Earth Data Initiative (BEDI)

Data DOIs

CERES Production Software Overview

Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	1	9K		
1	Instrument	5 + lib	157K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	11	358K		
4.5 – 4.6	Inversion	11	110K	SSF	1/hour
5	SARB	1	36K		1/hour
6 & 9	TISA-Gridding	7	34K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		
7.1 & 8 10	TISA-Averaging	4	102K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/month, 1/month, 5/month
12	MOA	2	14K		
	CERESlib/Perl_Lib		131K		
	Total	53	1,070K		



Current Data Availability (Edition 3)

Product	Platform	Processed through	Publically Available
BDS (Ed 3)	Terra & Aqua	June 30, 2014	Yes
SSF (Ed 3)	Terra & Aqua	June 30, 2014 (FM1 & FM3 only)	Yes
SFC (Ed 3)	Terra & Aqua	June 2014	Yes
SYN1deg (Ed 3)	Merged	May 2014	Yes
ISCCP-D2like-Day/Nit (Ed3)	Terra Aqua	June 2014 June 2014	Yes
ISCCP-D2like-GEO (Ed3)	GEO	June 2014	Yes
ISCCP-D2like-Mrg (Ed3)	Terra + Aqua + GEO	June 2014	Yes



Current Data Availability (Edition 4 & Edition 1 – NPP)

Product	Platform	Processed through	Publically Available
BDS (Ed 1)	NPP	January 2013	Yes
SSF (Ed 1)	NPP	February 2012	Yes
BDS (Ed 4)	Terra & Aqua	December 2011	No
SSF Beta2 Ed4	Terra Aqua	November 30, 2007 December 31, 2007	No
SSF Edition 1-CV	Terra Aqua	July 31, 2014	No
SSF (Ed4)	Terra Aqua	Reprocessing To Start This Week	No



Recent Activity

23 Software & Data Deliveries Since April 22nd, 2014

- CATALYST (5)
 - Post go-live fixes & performance improvements
- PR Tool (4)
 - Added new PGEs & site layout improvements
- Instrument (2)
- Clouds (1)
- Inversion (3)
- SARB (1)
- TISA Gridding (3)
- TISA Averaging (2)
- Perl_Lib (1)
 - Added improvements to PCF library
- CERESlib (1)
 - Added support for Edition4 TISA products



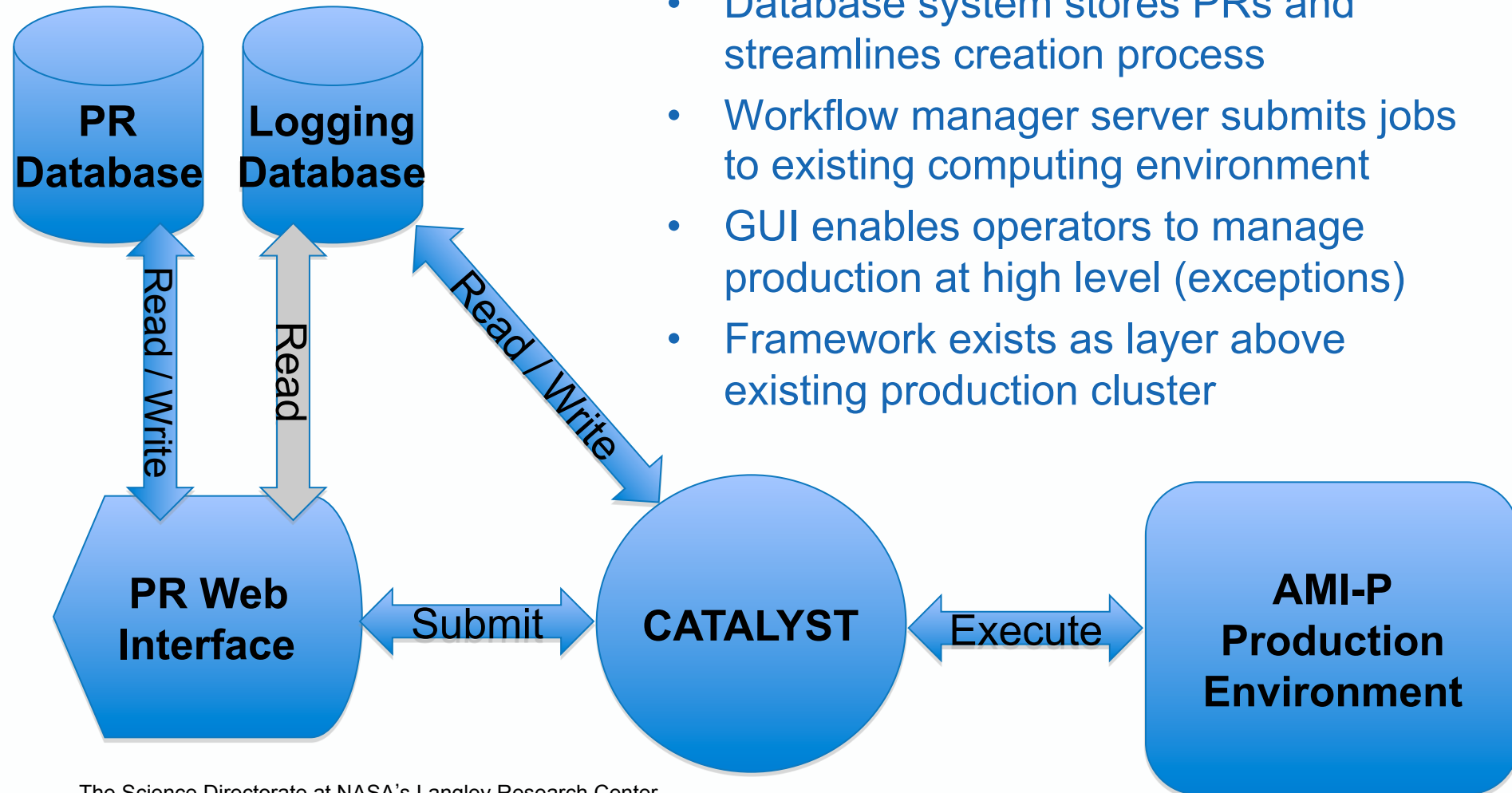
Production Automation Effort



Automation Framework

CERES AuTomAted job Loading sYSTem (CATALYST)

- Database system stores PRs and streamlines creation process
- Workflow manager server submits jobs to existing computing environment
- GUI enables operators to manage production at high level (exceptions)
- Framework exists as layer above existing production cluster





CATALYST Progress and Overview

CATALYST Development Team NASA Group Achievement Award – October 6th !!

CATALYST Build 1.0.4 started operating in production April 16th, 2014

- First 3.5 weeks in production saw sustained throughput rates that matched previous max short-term bursts (processing Beta1Ed4 SSF with E.E.W)
- Archive system failures forced CATALYST production pause: Running too fast!

Build 1.0 patches (5 total):

- Server v1.0.5 – Delivered in April and Promoted in May
 - Need and details identified prior to Go-Live
- Server v1.0.6 – Delivered in June and promoted early July
 - Urgent need identified for permissions updates and logging epilogue activities
- Operators' Console v1.0.5 – Delivered & Promoted August
 - Correct timeout bug
- Operators' Console v1.0.6 – Delivered & Promoted August
 - Correct additional console bug
- Server Configuration change (no code delivery) delivered Late August and promoted Early September
 - Changed CATALYST server to read files from local directory



CATALYST Release 2.0

CATALYST Release 2.0

- Addresses lessons learned from build 1.0
- Primarily resolves known and potential issues by reworking some server code to be more robust and maintainable
- Implements some new functionality where identified as needed by team
- Release 2.0 is decision gate to add remaining CERES subsystems to into CATALYST

Stand alone monitoring utility delivery (Release 2.0.1)

- Provide increased functionality for system health monitoring utility
- Respond to data center operator feedback to update existing monitoring utility



CATALYST Planned Releases

After Release 2.0, subsequent deliveries target PGE additions

Release 3: Add Edition 4 SSF1deg stream

- Add Edition 4 Inversion PGE logic
- Add Edition 4 SSF1deg-Hr PGE logic
- Add Edition 4 SSF1deg-Day/Month logic

Release 4: Edition 4 SYN1deg Stream

- Add Edition 4 TSI-> SYNI-> SYN1deg PGE logic

Release 5: Instrument and ERBElike processing

- Includes all processing for Terra, Aqua and NPP



CATALYST Planned Releases

Release 6: Edition 4 CRS and Edition 1 NPP CRS Stream

- Add all PGE logic for CRS PGEs

Release 7: ISCCP-D2Like & Flux-By-Cloud Streams

- Add ISCCP-D2Like PGE logic
- Add Flux-By-Cloud PGE logic



CATALYST Near-Term Schedule (Release 2.0)

Release 2.0 development and delivery

- Development effort complete February 11th (3 sprints)
- Delivery and Independent Testing February 17th – March 2nd

Release to Data Center SI&T Team – March 2nd

Production Integration Testing – March 3rd – March 23rd

Peer review (ORR) – Earliest Between March 24th & 27th

Go-Live Earliest Possible – March 30th, 2015

Release 3: SSF1deg stream – April 30, 2015



PR Tool Release 2

Initial version promoted to production March 1st, 2013

Release 2.0 to implement lessons learned and user feedback (Users internal to CERES Data Management and ASDC Teams)

- Series of updates to query capabilities
- Add tools to improve the PR creation and approval process – address existing deficiencies

Release 2 to be implemented as 3 increments

- Release 2.0 – Deliver November 21st, 2014
- Release 2.1 – Deliver February 27th, 2015
- Release 2.1 – Deliver April 24th, 2015



Planned Software Deliveries and Data Product Availabilities

Planned Delivery and Release Milestones (1 of 2)

Product	Science Delivery	Target Public Release
Edition 1 NPP BDS & SSF (Time varying Gains & SRFs)	Delivered	October 6 th , 2014
Ed4 Inversion	Delivered	October 30 th , 2014
Ed4 SSF-deg-Day/Month	October 17 th , 2014	January 15 th , 2015
Ed4 TSI code	December 5 th , 2104	March 30, 2015
Ed4 SYNI code	December 12 th , 2014	
Ed4 SYN1deg	December 19 th , 2014	



Planned Delivery and Release Milestones (2 of 2)

Product	Science Delivery	Target Public Release
Ed3 SSF1deg-Month	November 21 st , 2014	January 30 th , 2015
Ed4 ISSCP-D2like-Day/ Nit	January 30 th , 2015	April 15 th , 2015
Ed4 ISCCP-D2like-GEO	February 13 th , 2015	April 30 th , 2015
Ed4 ISCCP-D2like-Mrg	February 27 th , 2015	May 15 th , 2015
Ed4 Flux-By-Cloud-Type	March 27 th , 2015	June 15 th , 2015



MODIS C5 vs C6

MODIS Collection 5 soon to be replaced by Collection 6

Disk space considerations

- C5 & C6 ~7TB/Year
- Combined 371+TB not practical to keep on spinning disk

Planned Approach

- Move all C6 onto disk (~182TB)
- Keep only C5 seasonal months on disk (~59TB)

Result: 130TB disk space savings (241TB vs 371TB combined total)



System Migration Update

Currently nearly all CERES production code runs on AMI-P (IBM x86 and P6/P7 architectures) dedicated system

Exception = Edition 1-CV Clouds and Inversion on Magneto P4

- Inversion code has been delivered to AMI-P
- Clouds code initial migration encountered difficulties minimizing cross platform differences therefore decided not to migrate & Keep on legacy system

Legacy IBM P4 compute cluster beyond design life but reliable to date

- Experience shows failures typically limited to compute nodes and individual disks (as long as consistently powered on)

December data center full power outage (first in 3+ years)

- Anticipate high risk for significant disk system failure at power on

Will likely need to revisit Clouds code migration



Atmospheric Sciences Data Center (ASDC) Status Update



Upcoming ASDC Outages

Fall Semi-Annual Maintenance Cycle – October 17th to 24th

- Friday, October 17th
 - UGE submit queues to be taken offline @ 8am
 - Several disk arrays to go offline for routine file system checks (expected to run through Tuesday, October 21st)
- Tuesday & Wednesday, October 21st – 22nd
 - Main network switch down for maintenance
 - Most cluster-based systems unavailable
- Thursday, October 23rd (NLT)
 - All systems return to operational status

Full Power Outage – December 12th to 19th

- First “fully dark” power outage at ASDC in several years
- Power Infrastructure Improvement work
 - New HVAC and Power Distribution Unit (PDU) installation
 - Diesel generator maintenance
- Don't expect another power outage for several years
- Concurrent with AGU



ASDC, CERES and the Big Earth Data Initiative (BEDI)



Big Earth Data Initiative (BEDI)

2014 President's budget request included funds to kick off BEDI with \$9m each for NOAA and USGS, \$6M for NASA, and \$4M for USDA.

BEDI includes practices and recommendations that improve the Discoverability, Accessibility and Usability of Earth Science data sets.

The ASDC is asking for guidance from CERES to develop a list of the most useful L2 and L3 datasets and specific parameters to support BEDI.

- An initial dataset list includes EBAF, SSF and SYN1deg
- Data set parameters to be available via OpenDAP and Web Map Service (WMS) for WMS application access (ArcView etc)

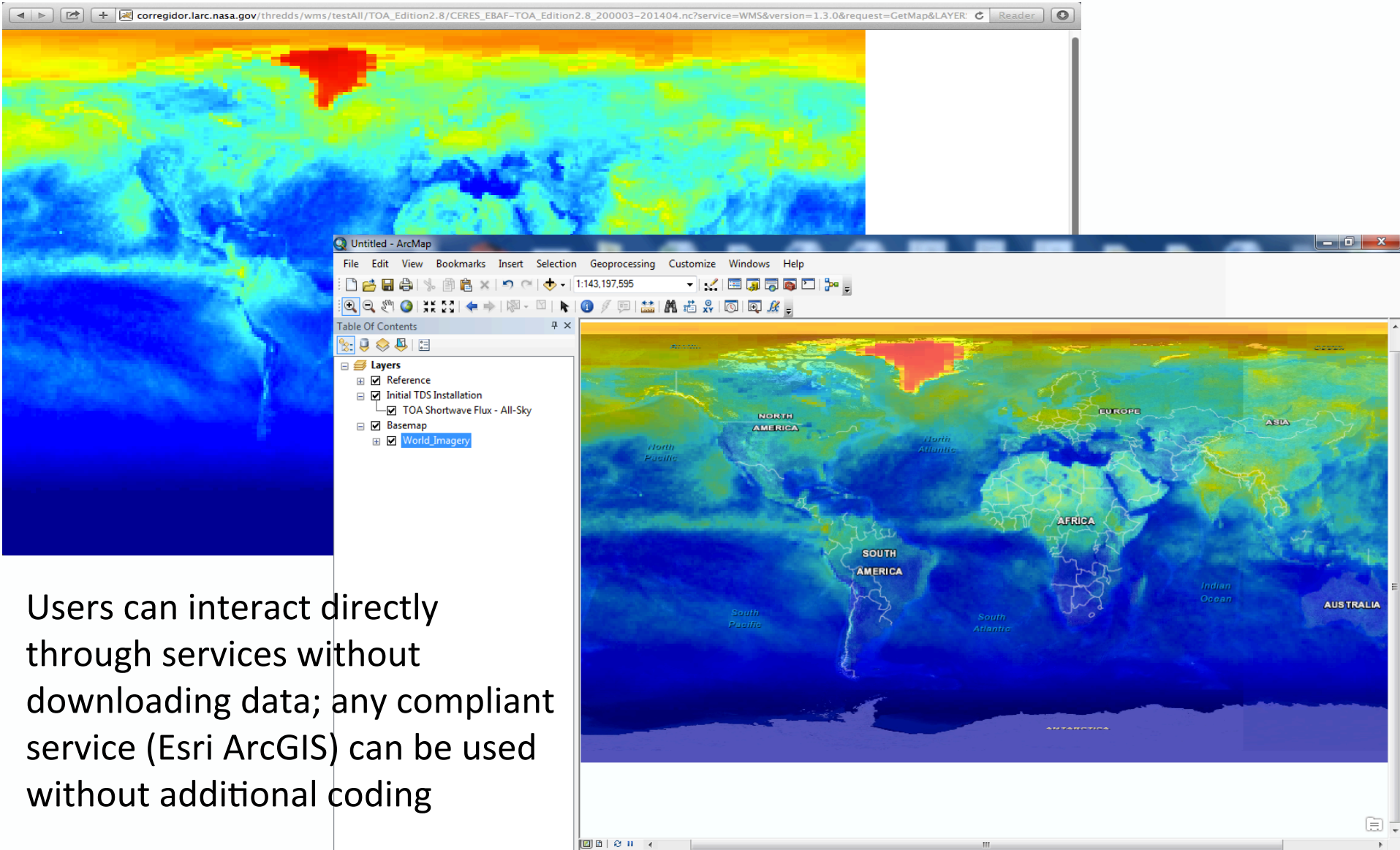
Dataset: TOA_Edition2.8/CERES_EBAF-TOA_Edition2.8_200003-201404.nc

- Access:**

- THREDDS OPeNDAP server exposes HTTP service to download EBAF TOA Ed 2.8 in NetCDF; WMS maps the image and WCS can pass values to a model



Prototype CERES BEDI Service Access



Users can interact directly through services without downloading data; any compliant service (Esri ArcGIS) can be used without additional coding



Data Digital Object Identifiers (DOIs)

Secondary objective of BEDI to implement DOIs for all EOS datasets

ASDC initiative currently planned to assign DOIs to all data center holdings

For CERES will start with Edition 4 SSF1deg and SYN1deg products

- Can be completely unobtrusive
- Since starting with Ed4 Tisa, will record the DOI for the product in the .met file and the HDF file
- “Data home pages” already provided by ASDC

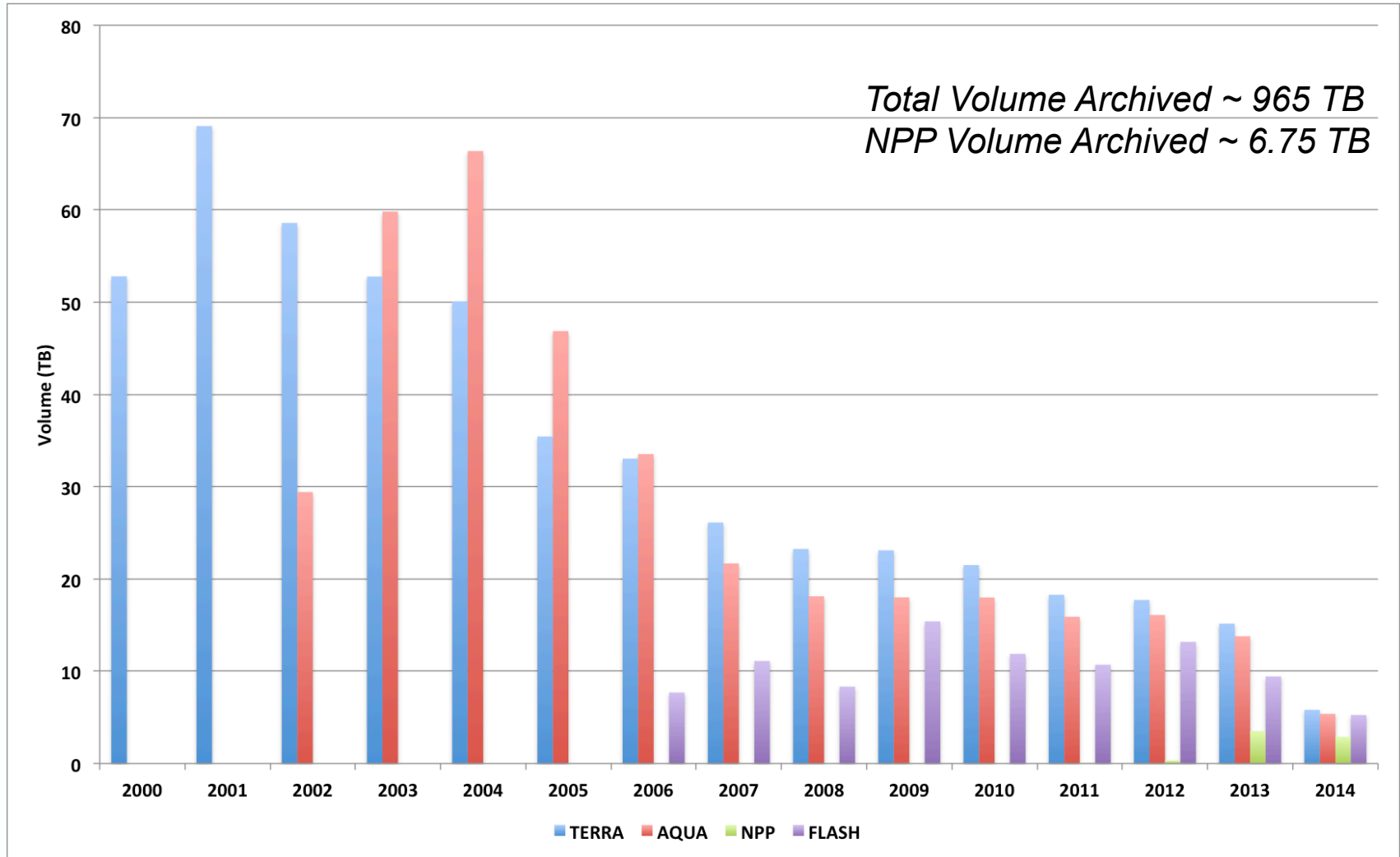
All DOI tagged datasets registered at doi.org

- Provides easy access to data set referencing information for publisher and CERES team



CERES and FLASHFlux Archive Volume

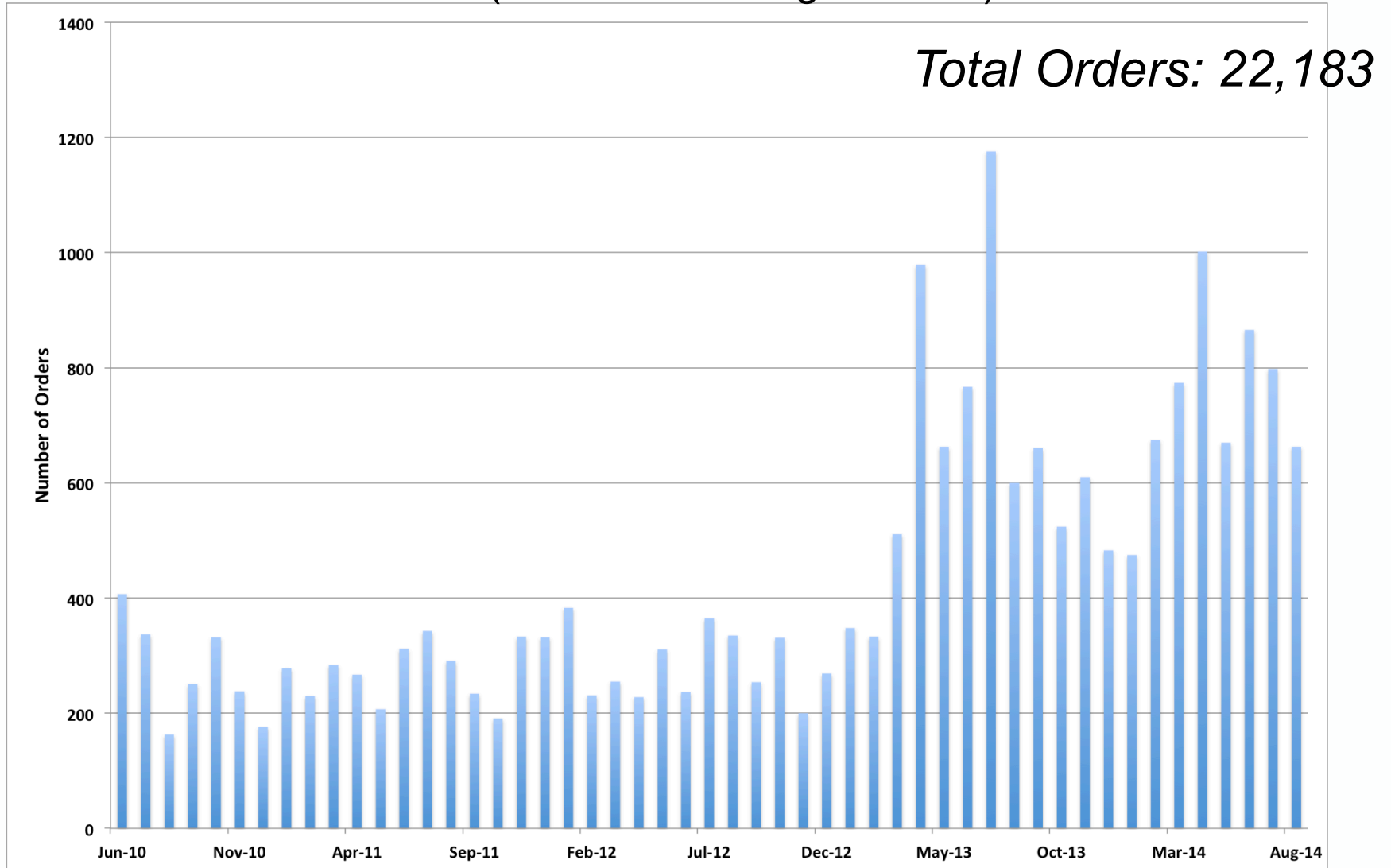
By Data Date Through August 2014





CERES Data Orders

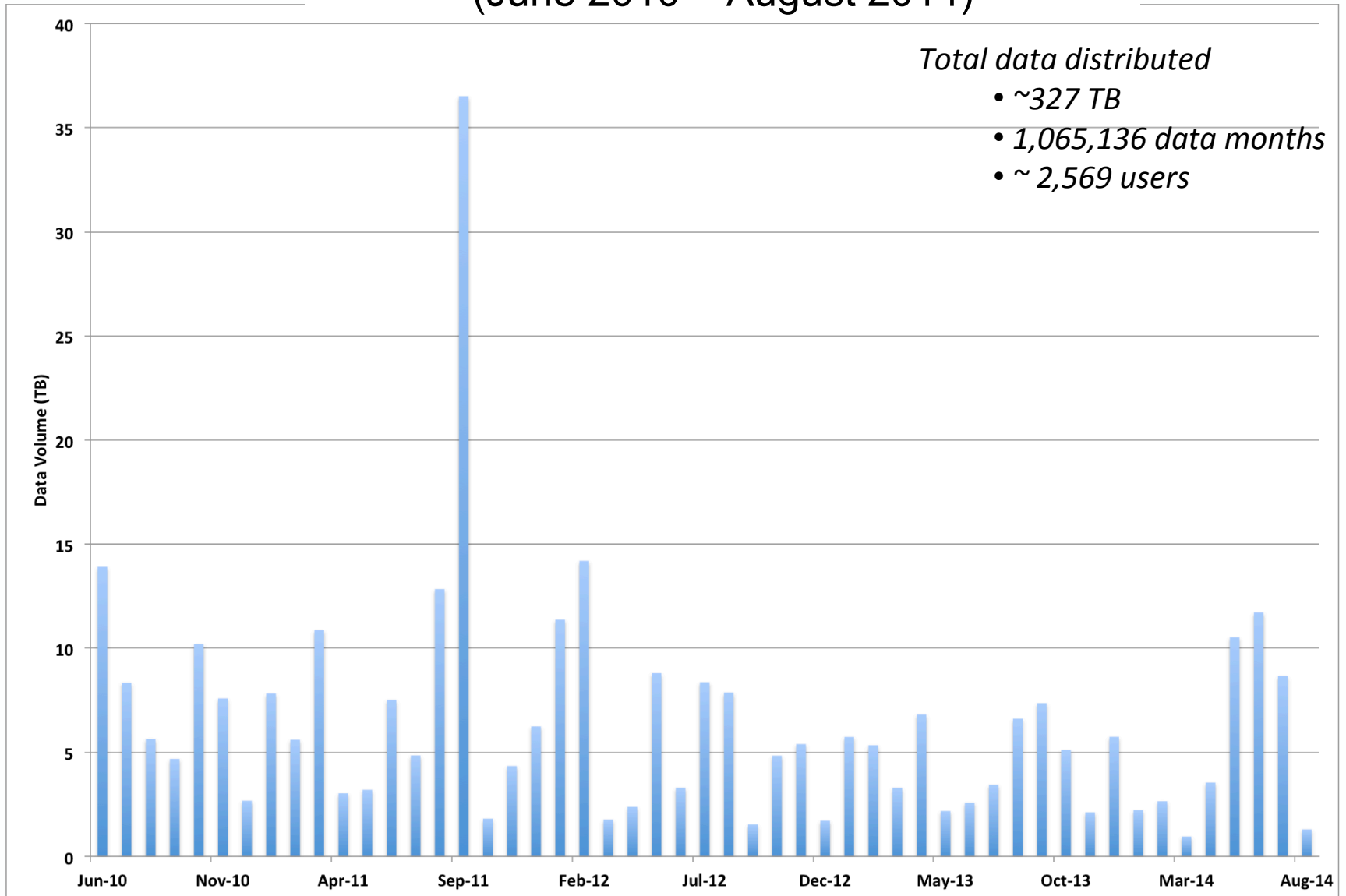
(June 2010 – August 2014)



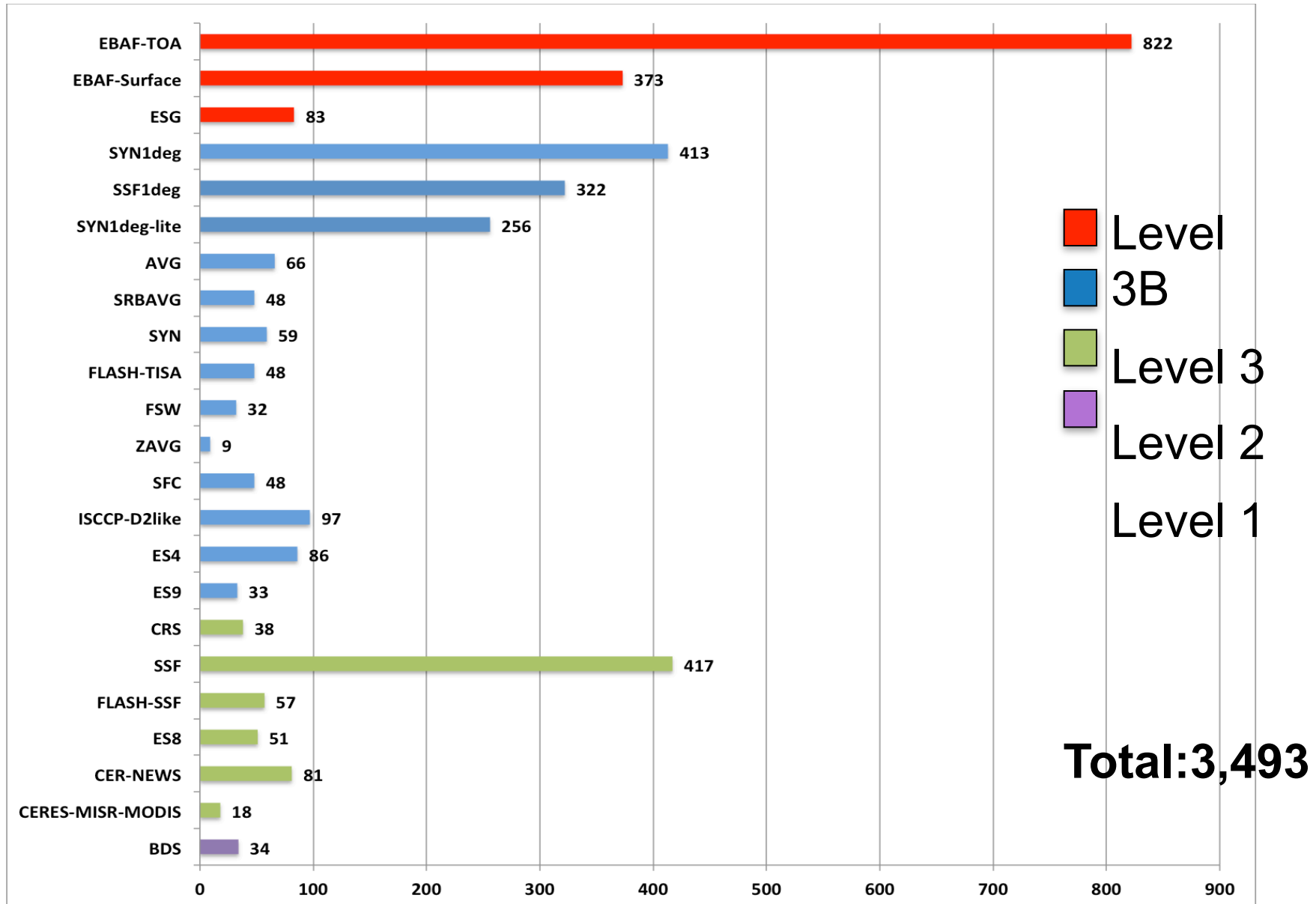


CERES Data Distribution

(June 2010 – August 2014)

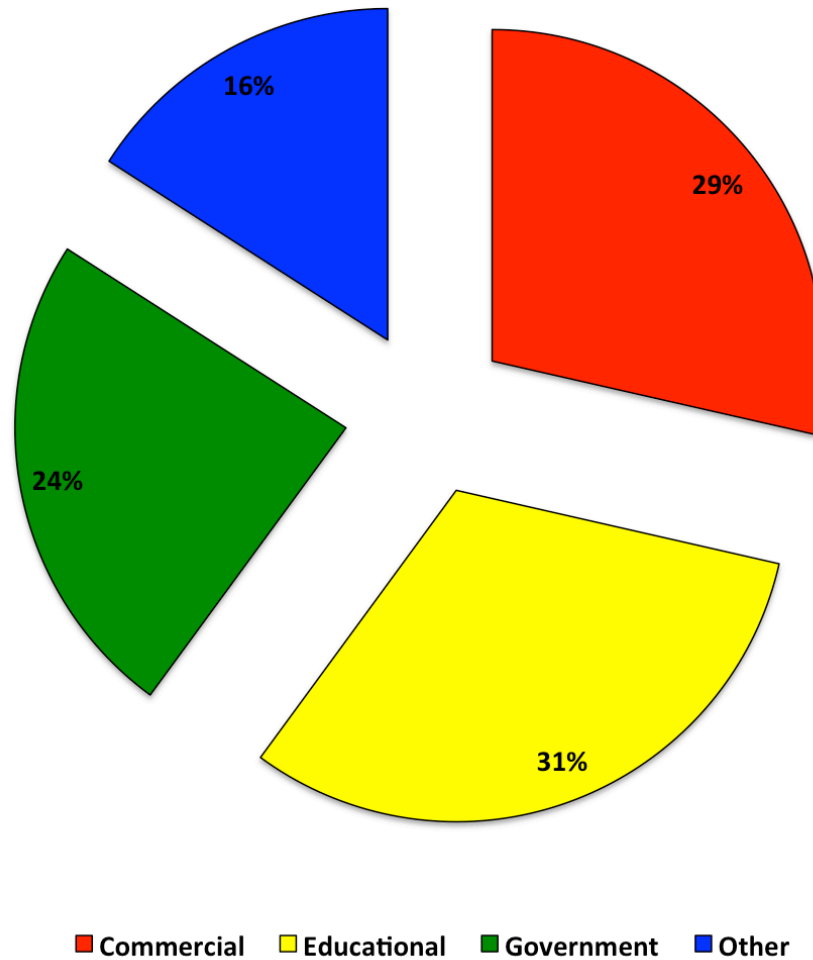


Number of Users by Product *(June 2010 – August 2014)*

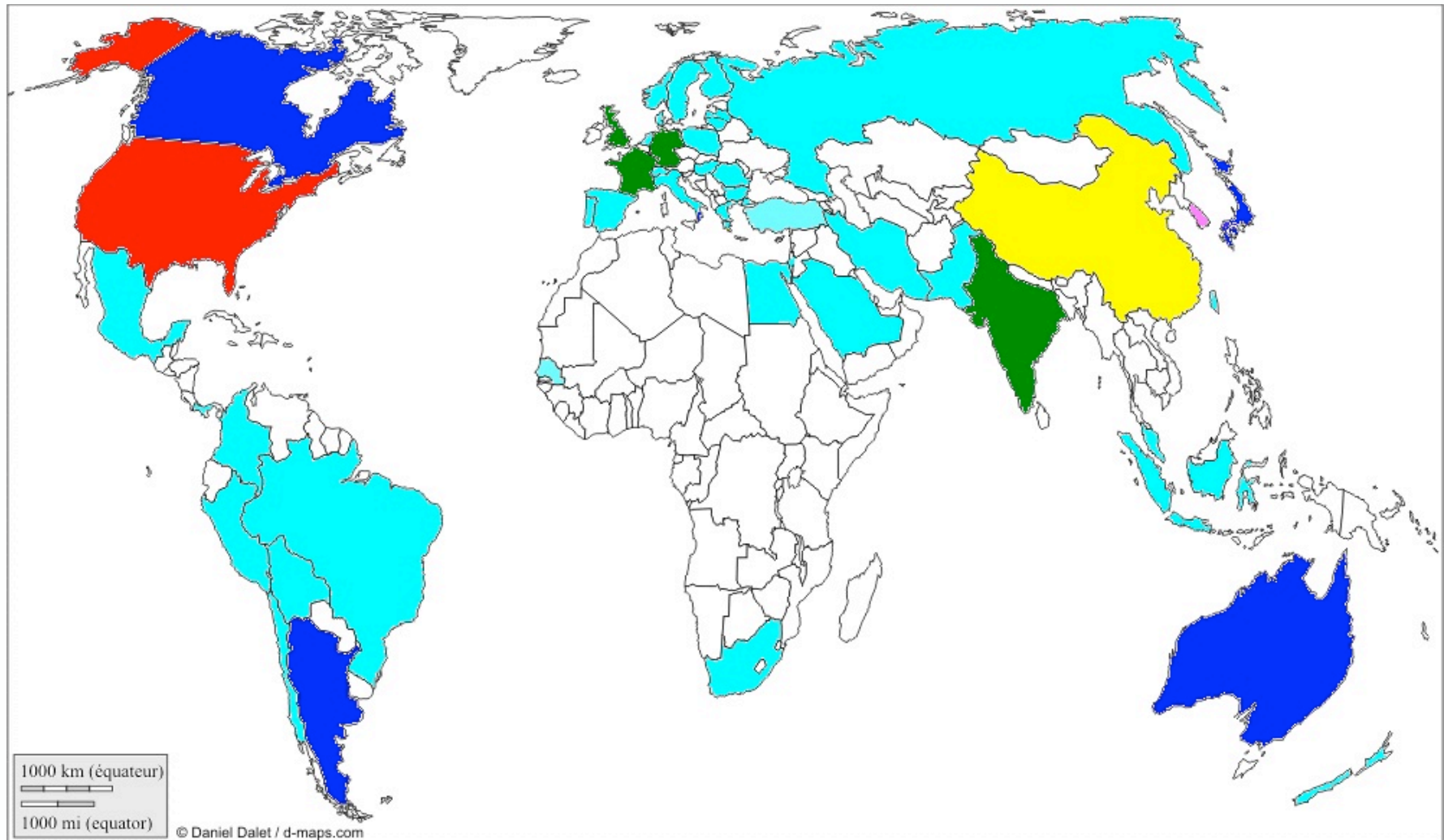




User Affiliations



Users by Country (June 2010 – August 2014)





Questions?



Backup Slides

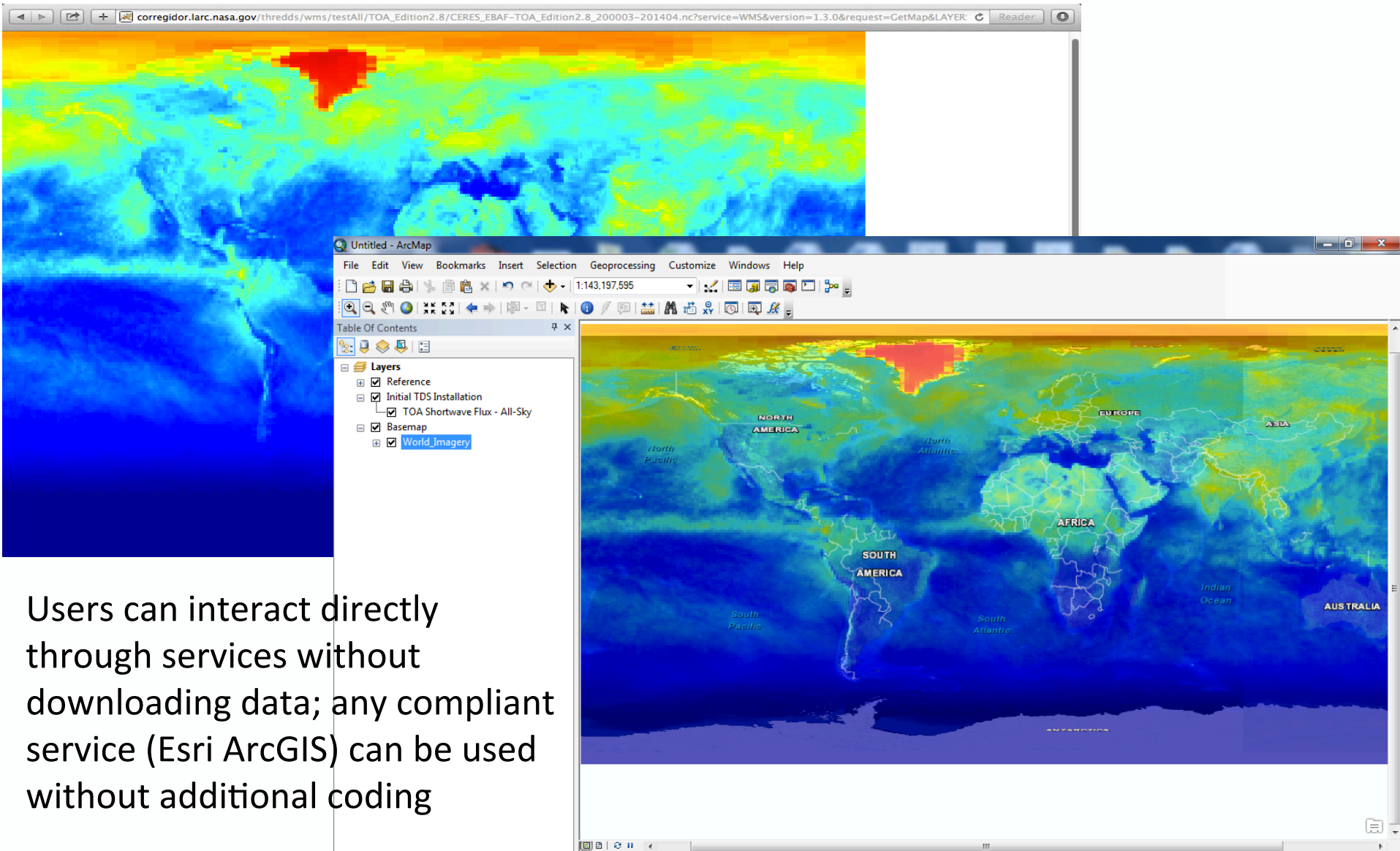
Dataset: TOA_Edition2.8/CERES_EBAF-TOA_Edition2.8_200003-201404.nc

- Access:**

- THREDDS OPeNDAP server exposes HTTP service to download EBAF TOA Ed 2.8 in NetCDF; WMS maps the image and WCS can pass values to a model



Prototype CERES BEDI Service Access



Users can interact directly through services without downloading data; any compliant service (Esri ArcGIS) can be used without additional coding



Motivation To Automate

- **Multi-year, resource intensive effort**
 - 2 + Years just go get to this point
 - PR Tool live X Months – Growing pains continue
 - CATALYST Go-Live =
- **Some NASA DAAC processing resources not fully utilized, ASDC included until recently.**
 - Can we then justify spending resources to automated?
 - Yes, Human In The Loop
- **Bursty reprocessing activities not consistent**
- **Big Iron Blues – Can't sustain CERES with windfall, end of FY funding.**
 - Initial Purchase + 3-5 year maintenance and then start over
 - Requires consistent, big funding
- **“Linux OS isn't as robust as the old mainframe OSs, They had decades to refine and just couldn't afford to**



CATALYST

- **CERES AuTomAteD job Loading sYSTem (CATALYST)**
- Automation Framework for PGE Execution, Coordination & Logging that:
- Modularly implements science software dependency logic
- Leverages existing CERES production infrastructure
 - Ingest CERES PR to create collection of jobs
 - Build all possible individual jobs
 - Identify job input dependencies
 - Execute on AMI-P cluster
 - Send jobs to Sun Grid Engine
 - Utilize existing job submission scripts
 - Initiate ANGe ingest wrapper scripts
- Provides graphical interface for users to manage system



CATALYST (Cont'd)

- Automation Framework for PGE Execution, Coordination & Logging that:
 - Dynamically manages job input dependencies and determine when jobs are runnable
 - CATALYST jobs wait for predecessor jobs to complete
 - Predecessor jobs can be:
 - Internal to CATALYST
 - External to CATALYST (via a backlogging interface)
 - Jobs broadcast completion status to follow-on jobs for rapid follow-on execution
 - Stores job execution state long-term in a job logging database



CATALYST Implementation

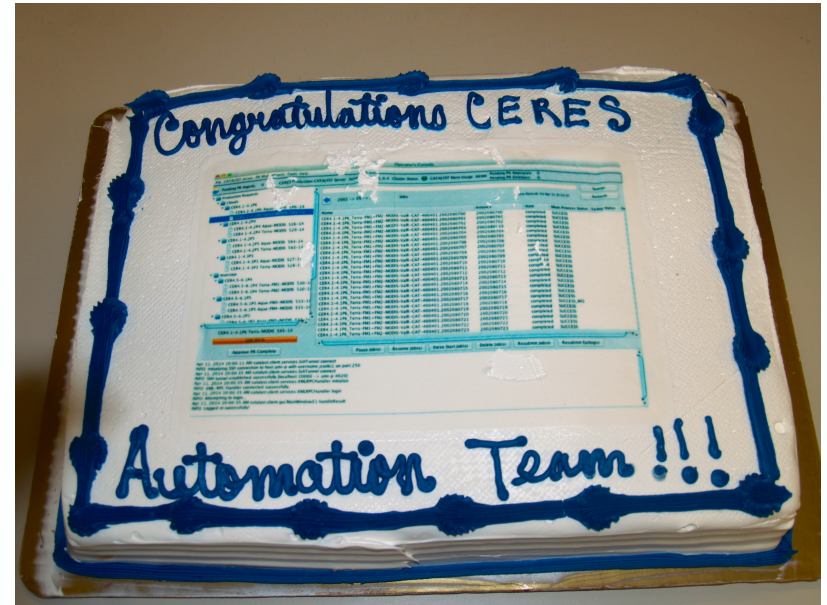
- Primary languages Perl & Java
 - Perl ~12.2K LOC
 - Java ~6.6K LOC
- Defines an Application Programming Interface (API) using externally accessible XML-RPC
 - Allows users to inspect and modify job execution status programmatically in any language (with XML-RPC libraries)
 - Enables development of modular tools to function in concert with CATALYST – backlogging component
- Uses existing AMI LDAP service for user authentication
 - Permissions defined in access control list



Timeline of Events

- .January 2012 – ASDC & DMT workshop**
- .3/1/2013 – PR Tool went “Live”**
- .3/22/2013 – CATALYST Test Readiness Review**
- .5/1/2013 – Implemented automation software Change Review Board**
- .11/19/2013 – Release current CATALYST Server v1.0.4**
- .12/9/2013 – Release current Operator Console v1.0.4**
- .12/19/2014 – CATALYST Operational Readiness Review**
- .3/18/2014 – Rebuild and update Logging DB (reflect DPO post recovery)**
- .3/18/2014 – Release standalone status monitoring & recovery utilities**
- .4/16/2014 – CATALYST Go-Live**

CATALYST Go Live!



The Science Directorate at NASA's Langley Research Center



CATALYST Release Schedule

•Release 1:

- Build 1.0.4 – Currently in production
- Build 1.0.5 - Patch delivery
 - Deliver for testing on 4/24/2014
 - Promote to production on 5/27/2014
- Build 1.0.6 – Patch delivery
 - Deliver for testing on 7/11/2014
 - Promote to production on 8/24/2014



CATALYST Development & Delivery Schedule Summary View



4/22/14

Key Milestones	FY14				FY15			
	1Q'14	2Q'14	3Q'14	4Q'14	1Q'15	2Q'15	3Q'15	4Q'15
Release 1: CATALYST Development & Deployment								
Operational Readiness Review (ORR)		▲ 12/19						
Run & Validate ValRs			4/8 4/10					
Promote to Production			▲ 4/16					
Build 1.0.5								
Development, Internal Testing, and Delivery to CM		3/24	■ 4/24					
Operational Testing			4/25 ■ 5/9					
ANGe Outage			5/12 ■ 5/16					
Complete Operational Testing/Promote to Production			5/19 ▲ 5/27					
Build 1.0.6								
Development, Internal Testing, and Delivery to CM			5/28 ■ 7/11					
Operational Testing/Promote to Production			7/14 ■ 8/4					
Release 2: Edition 4 SSF-1deg Stream								
Development, Internal Testing, and Delivery to CM				8/5	■ 10/15			
Operational Testing/Promote to Production					10/15	■ 11/27		
Release 3: Edition 4 SYN1deg Stream								
Development, Internal Testing, and Delivery to CM					11/27	■ 1/12		
Operational Testing/Promote to Production						1/13	■ 2/3	
Release 4: Edition 3/4 Instrument & ERBElike Stream								
Development, Internal Testing, and Delivery to CM						2/4	■ 3/6	
Operational Testing/Promote to Production						3/6	■ 3/26	
Release 5: Edition 1-CV/Baseline1-QC Instrument Stream								
Development, Internal Testing, and Delivery to CM						3/26	■ 4/27	
Operational Testing/Promote to Production							4/28	■ 5/12
Release 6: Edition 4/NPP CRS								
Development, Internal Testing, and Delivery to CM							5/13	■ 6/1
Operational Testing/Promote to Production							6/2	■ 6/19
Release 7: Edition 4 ISCCP & Flux-By-Cloud								
Development, Internal Testing, and Delivery to CM							6/22	■ 7/23
Operational Testing/Promote to Production							7/24	■ 8/12

■ Work Effort

△ Open Milestone

▲ Completed Milestone

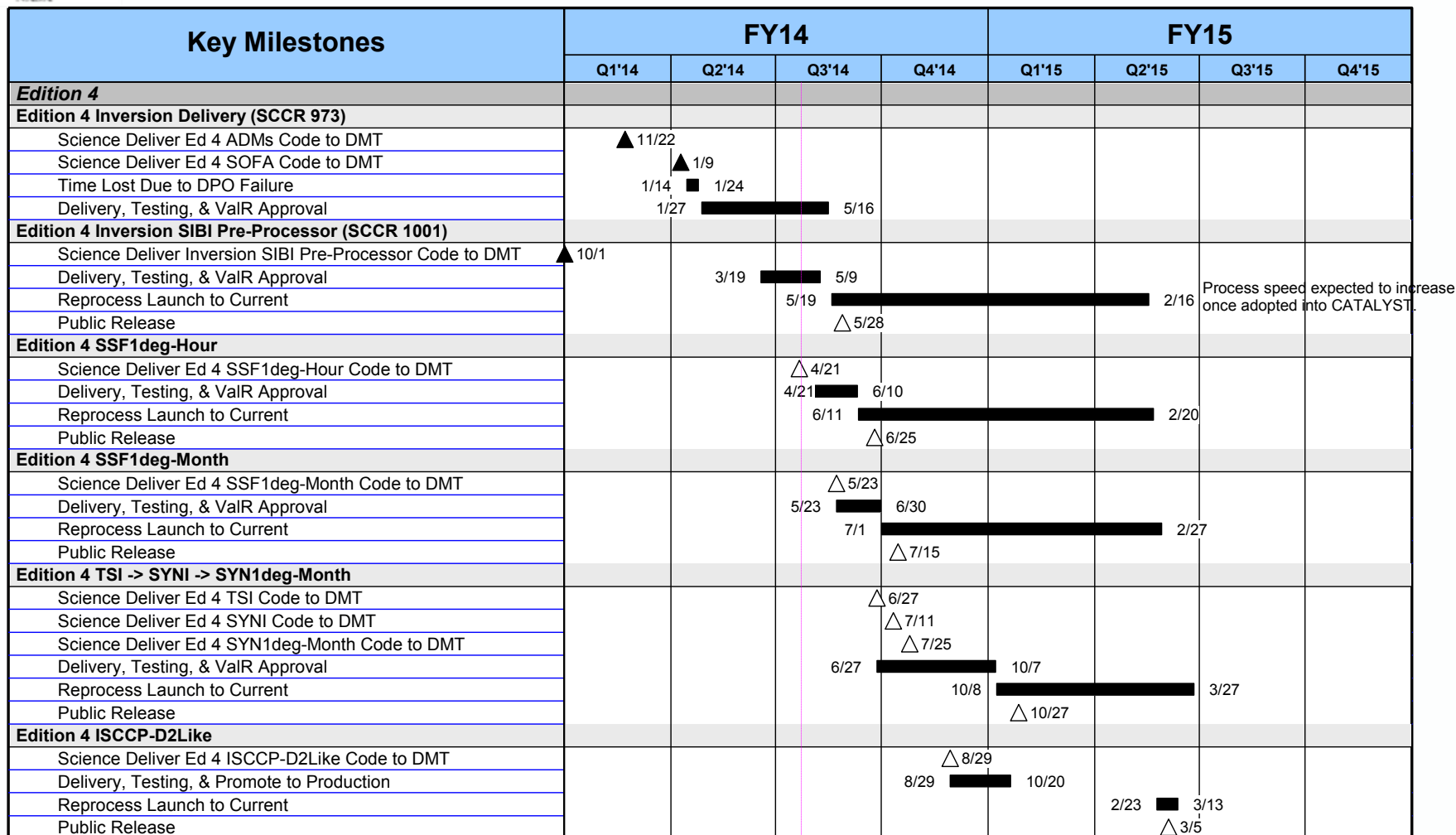


CERES Software Delivery and Data Processing Schedule

Edition 4, NPP, & Edition 3



4/22/14



△ Open Milestone

▲ Completed Milestone

■ Work Effort



CERES Software Delivery and Data Processing Schedule

Edition 4, NPP, & Edition 3



4/22/14

Key Milestones	FY14				FY15			
	Q1'14	Q2'14	Q3'14	Q4'14	Q1'15	Q2'15	Q3'15	Q4'15
Edition 4 Flux-by-Cloud Type								
Science Deliver Ed 4 Flux-by-Cloud Type Code to DMT					△ 10/31			
Delivery, Testing, & ValR Approval					10/31 ■ 12/18			
Reprocess Launch to Current					12/19 ■ 1/16			
Public Release					△ 12/31			
NPP								
NPP Edition 1 SSF								
Science Deliver Ed 1 Clouds Code to DMT		▲ 2/14						
DMT Prepare Delivery Package Using Ed4 ADM Code		4/7 ■ 5/9						
Delivery, Testing, & ValR Approval		2/18 ■ 7/1						
Reprocess Launch to Current			7/2 ■ 9/18					
Public Release					△ 9/26			
Produce Edition 1 IES with Time Varying Gains								
Instrument Group Deliver NPP Time Varying Gains								
Reprocess Instrument (via 1.4P3)								
Publicly Release NPP BDS with Time Varying Gains								
Optional Redelivery of NPP Inversion-Only Code								
Science Deliver Edition 1 SOFA Code to DMT					△ 10/10			
NPP Edition 1 CRS (SCCR 975)								
Science Deliver Ed 1 CRS Code to DMT					△ 11/14			
Delivery, Testing, & ValR Approval					11/14 ■ 1/5			
Reprocess Launch to Current					1/6 ■ 1/12			
Public Release					△ 1/14			
Edition 3								
Edition 3 SSF1-deg-Day/Month								
Science Deliver Ed 3 SSF1deg-Month Code to DMT								
Delivery, Testing, & ValR Approval								
Reprocess Launch to Current								
Public Release								

△ Open Milestone

▲ Completed Milestone

■ Work Effort